

Versión en español


<https://aplicaciones.uc3m.es/est=340&anio=2021&plan=376>

Course: 2021/2022

Information retrieval systems (17449)

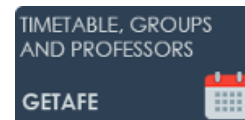
Study: Bachelor in Management of Information and Digital Contents (340)

Coordinating teacher: SERRANO LOPEZ, ANTONIO ELEAZAR

Department assigned to the subject: Department of Library Science and Documentation

Type: Compulsory ECTS Credits: 6.0 ECTS

Course: 2º Semester: 1º


<https://aplicaciones.uc3m.es/consultaHorarios/porAsignatura.htm?ano=2021¢ro=3&plan=376&asignatura=17449&idioma=en>

<https://aplicaciones.uc3m.es/cpa/cpa/generaFichaPDF?ano=2021&plan=376&asignatura=17449&idioma=2> () ()

Requirements (Subjects that are assumed to be known)

Any

Objectives ()

In general, search and retrieve information, thanks to computer and manual methods and tools, that allow responding to the demands of users in optimal conditions of costs and terms, and evaluate the adequacy between the demand and the response provided, as well as determining and to evaluate the technological needs related to the management of documentary databases that may be of interest and useful at the present time or in the near future for the services and information units. And specifically:

1. Understand and know the definition of terms related to Information Retrieval (RI).
2. Understand and apply the principles and techniques for IR and its evaluation.
3. Know the theoretical models of information retrieval.
4. Handle with ease the different interrogation languages and interfaces of information retrieval systems.
5. Interact with the information retrieval systems to solve the possible information needs that may arise.
6. Distinguish the different theoretical models of IR and recognize them in the Real Information Retrieval Systems (IRS).
7. Convert a request for information into a search strategy appropriate to the system and transcribe and transmit the results of a search.
8. Handle with ease, compare and evaluate different interrogation languages and interfaces that allow interaction with a local SRI or with engines, metasearch engines and other search tools in the network.
9. Master at least one software for information retrieval, advanced features, extended installation and recognized quality, which serves as a basis for the analysis and evaluation of any other.
10. Evaluate the results of a search in terms of reliability and relevance, in any environment of use of an SRI.

Skills and learning outcomes ()

[Link to document \(https://www3.uc3m.es/reina/RESULTADOS_APRENDIZAJE/Idioma_1/2021/340.17449.pdf\)](https://www3.uc3m.es/reina/RESULTADOS_APRENDIZAJE/Idioma_1/2021/340.17449.pdf)

Description of contents: programme

THEORETICAL CONTENT

LEARNING UNIT 1: Introduction to Information Retrieval Systems (IRS).

- Lesson 0: Information Retrieval (IR) in text databases
- Lesson T1: The theoretical framework for IR: Relation to Indexing; difference with Data Retrieval

LEARNING UNIT 2: Main formal models of IR (as D. Blair).

- Lesson T2: Basic models (Models 1-4): one descriptor, several descriptors, cutoff value and ordered output
- Lesson T3: Models with weighted descriptors (Models 5-8): weighted search only, only weighted indexing, weighted search and indexing and search in a vector space
- Lesson T4: Models with Boolean search (Models 9 and 10): Boolean search and peculiarities on free text
- Lesson T5: thesaurus-based models (Models 11 and 12): Search with binary and weighted thesauri

LEARNING UNIT 3: Evaluation of Information Retrieval Systems.

- Lesson T6: Principles for evaluating the effectiveness of retrieval: Relevance, Recall and Precision and its complements; relationship between P and R; other measures
- Lesson T7: Difficulties in obtaining indicators: The problem of Silence; Relevance, as Affinity (relatedness) and Utility; Relevance, as binary or weighted

PRACTICAL CONTENT

LEARNING UNIT 4: Analysis of Information Retrieval software: BKM (based on BRS).

- Lesson P1: BKM (1). Organisation of databases; interfaces and basic functions; Boolean operators; search criteria and strategy fractionation; reference to previous sets; proximity, desired and avoided, operators; priorities of operators
- Lesson P2: BKM (2). Limitation to fields, searched and ignored; truncation and wildcards; using the indexes; History and archiving of strategies; display and export formats
- Lesson P3: BKM (3). Selection of work universe; numerical search; sorting and exporting results; search through the integrated thesaurus, active and passive; user preferences

LEARNING UNIT 5: IR in online distributed databases environment.

- Lesson P4: Entrez (1). Special features of IR in this environment: search preprocessing; importance of control in terminology
- Lesson P5: Entrez (2). Using filters.
- Lesson P6: Solr/Lucene (1). Special IR features in this environment: Faceted search (or navigation). Flexible relevance
- Lesson P7: Solr/Lucene (2). Search results clustering. Appendix: Comparison of these environments with internet search engines (Google and others). Special features of IR in this environment: poor structure and little control in terminology; importance of results order; Precision vs. Recall

Learning activities and methodology

- Acquisition of theoretical knowledge (total 3 credits ECTS) through lectures, teaching materials prepared by the teacher, online tutorials, specialized readings and discussions (1.2 ECTS), and the personal study and work of students (1.8 ECTS). It relates to the abilities 1 to 3.

- Acquisition of practical skills (total 3 ECTS) through various practical assumptions of information retrieval in different environments (local systems, online and other websites), with which they can acquire skills and develop abilities 4 to 10.

- Tutorship: The schedule of tutorship sessions could be looked up in the Aula Global space for the course. In addition to the tutorship at the times and places officially set for the course, students can apply for other outside these hours and to be held by digital media.

Assessment System

The final course grade will be determined by the following factors and their corresponding weights:

1. Continuous assessment [= 60% of the final score, 6 points out of 10]

- Tests: 30%, 3 points out 10

- Practicing: 30%, 3 points out 10

2. Final exam [= 40% of the final score, 4 points out 10]

According to University policy, in the regular exam session, the student who did not follow the continuous assessment is entitled to take an exam for the 60% of the final grade.

In the extra exam session, if the student did not follow the continuous assessment, is entitled to take an exam for the 100% of the final grade, and if he did follow the continuous assessment, will be evaluated in the most beneficial way, considering an exam weight of 40% plus the continuous assessment score, or an exam weight of 100%, discarding the score obtained in continuous assessment.

- % end-of-term-examination 60
- % of continuous assessment (assignments, laboratory, practicals...) 40

Calendar of Continuous assessment (https://www3.uc3m.es/reina/CALENDARIO_EVAL_CONTINUA/Idioma_1/2021/340.17449.pdf)

Basic Bibliography

- BLAIR, D.C.. Language and Representation in Information Retrieval.. Elsevier Science Publishers. 1990
- Baeza-Yates, R.; Ribeiro-Neto, B. Modern information retrieval. Addison-Wesley. 1999
- CHOWDHURY, G.G.. Introduction to modern information retrieval (3ª ed.). Library Association. 2010
- LANCASTER, F.W. El control del vocabulario en la recuperación de la información (2ª ed. corr.). Universitat de València. 2002
- MEADOW, CH.T.; BOYCE, B.R.; KRAFT, D.H. Text information retrieval systems (3ª ed.). San Diego, Academic Press. 2007

Additional Bibliography

- Buckland, M.K.. Information and Information Systems. Greenwood Pres. 1991
- Chamis, A.Y.. Vocabulary Control and Search Strategies in Online Searching. Greenwood Press. 1991
- Manning, C.D.; Raghavan, P.; Schütze, H.. Introduction to Information Retrieval.. Cambridge University Press. 2008
- Meadow, Ch.T. Text information retrieval systems. Academic Press. 2000
- Salton, G. Automatic text processing: The Transformation, Analysis, and Retrieval of Information by Computer. Addison Wesl. 1989

The course syllabus and the academic weekly planning may change due academic events or other reasons.